



Haitian Police Station Video

2023-21-FB-MR

Summary

The Oversight Board has overturned Meta’s decision to take down a video from Facebook showing people entering a police station in Haiti, attempting to break into a cell holding an alleged gang member and threatening them with violence. The Board finds the video did violate the company’s Violence and Incitement policy.

Nonetheless, the majority of the Board disagrees with Meta’s assessment on the application of the newsworthiness allowance in this case. For the majority, Meta’s near three-week delay in removing the content meant the risk of offline harm had diminished sufficiently for a newsworthiness allowance to be applied. Moreover, the Board recommends that Meta assess the effectiveness and timeliness of its responses to content escalated through the Trusted Partner program.

About the Case

In May 2023, a Facebook user posted a video showing people in civilian clothing entering a police station, attempting to break into a cell holding a man – who is a suspected gang member, according to Meta – and shouting “we’re going to break the lock” and “they’re already dead.” Towards the end of the video, someone yells “bwa kale na boudaw,” which Meta interpreted as a call for the group to “to take action against the person ‘bwa kale style’ – in other words, to lynch him.” Meta also interpreted “bwa kale” as a reference to the civilian movement in Haiti that involves people taking justice into their own hands. The video is accompanied by a caption in Haitian Creole that includes the statement, “the police cannot do anything.” The post was viewed more than 500,000 times and the video around 200,000 times.

Haiti is experiencing unprecedented insecurity, with gangs taking control of territory and terrorizing the population. With police unable to address the violence and, in some instances, said to be complicit, a movement has emerged that has seen “more



than 350 people [being] lynched by local people and vigilante groups” in a four-month period this year, according to the UN High Commissioner for Human Rights. In retaliation, gangs have taken revenge on those believed to be in or sympathetic to the movement.

A Trusted Partner flagged the video to Meta as potentially violating 11 days after it was posted, warning the content might incite further violence. Meta’s Trusted Partner program is a network of non-governmental organizations, humanitarian agencies and human rights researchers from 113 countries. Meta told the Board that the “greater the level of risk [of violence in a country], the higher the priority for developing relationships with Trusted Partners,” who can report content to the company. About eight days after the Trusted Partner’s report in this case, Meta determined the video included both a statement of intent to commit and a call for high severity violence and removed the content from Facebook. Meta referred this case to the Board to address the difficult moderation questions raised by content related to the “Bwa Kale” movement in Haiti. Meta did not apply the newsworthiness allowance because the company found the risk of harm was high and outweighed the public interest value of the post, noting the ongoing pattern of violent reprisals and killings in Haiti.

Key Findings

The Board finds the content did violate Facebook’s Violence and Incitement Community Standard because there was a credible threat of offline harm to the person in the cell as well as to others. However, the majority of the Board disagrees with Meta on the application of the newsworthiness allowance in this case. Given the delay of nearly three weeks between posting and enforcement, Meta should have applied the newsworthiness allowance to keep up the content, with the Board concluding the risk of harm and public interest involved in any newsworthiness analysis should be assessed at the time Meta is considering issuing any allowance, rather than at the time content is posted. The Board finds that Meta should update its language on the newsworthiness allowance to make this clear to users.

For the majority of Board Members, Meta’s near three-week delay in removing the content meant the risk of offline harm had diminished sufficiently for a newsworthiness allowance to be applied. This group considered the context in Haiti,



the extent and reach of the post, and the likelihood of harm given the delay in enforcement. By that time, when the video already had 200,000 views, the risk the content posed had already likely materialized. Furthermore, in a situation of protracted widespread violence and breakdown in public order, sharing information becomes even more important to allow communities to react to events, with the video holding the potential to inform people in both Haiti and abroad about the realities in the country.

However, a minority of Board Members find Meta was right not to apply the allowance. Since the content was posted during a period of heightened risk, the threat of the video leading to additional and retaliatory violence had not passed when Meta reviewed the content. These Board Members consider removal necessary to address these risks.

The Board is concerned about Meta's ability to moderate content in Haiti in a timely manner during this period of heightened risk. The delay in this case appears to be the result of the company's failure to invest adequate resources in moderating content in Haiti. Meta was not able to provide a timely assessment of the report from its Trusted Partner. Reports from Trusted Partners are one of the main tools Meta relies on in Haiti to identify potentially violating content. A recent report by a Trusted Partner found that Meta does not adequately resource its own teams to review content identified by Trusted Partners and there is significant irregularity in response times.

Finally, the Board notes Meta failed to activate its Crisis Policy Protocol in Haiti. While Meta told the Board it already had risk-mitigation measures in place, the Board is concerned the lengthy delay in this case indicates that existing measures are inadequate. If the company fails to use this protocol in such situations, it will not deliver timely or principled moderation, undermining the company's and the public's ability to assess the effectiveness of the protocol in meeting its aims.

The Oversight Board's Decision

The Oversight Board overturns Meta's decision to take down this content, requiring the post to be restored.

The Board recommends that Meta:



- Assess the timeliness and effectiveness of its responses to content escalated through the Trusted Partner Program, to address the risk of harm particularly where Meta has no or limited proactive moderation tools, processes or measures to identify and assess content.
- The Board also takes this opportunity to remind Meta of a previous recommendation, from the Russian Poem case, that calls for the company to make public an exception to its Violence and Incitement policy. This exception allows for content that “condemns or raises awareness of violence,” but Meta requires the user to make it clear they are posting the content for either of these two reasons.

*Case summaries provide an overview of the case and do not have precedential value.

Full Case Decision

1: Decision Summary

The Oversight Board overturns Meta’s decision to take down a Facebook post of a video depicting a group of people entering a police station in Haiti. As the crowd attempts to gain access to a locked cell holding an alleged gang member, members of the crowd shout, “we’re going to break the lock” and “they’re already dead,” and other phrases threatening violence. The Board finds the post did violate Meta’s Violence and Incitement policy as it depicts incitement to violence in a context where there is a credible threat of offline harm to the person in the cell as well as others. However, the majority of the Board disagrees with Meta’s assessment on the application of the newsworthiness allowance in this case. For the majority, given the near three-week delay in Meta removing the content, the risk of harm had significantly diminished, and Meta should have kept the content on the platform given the public interest value of the post. For a minority of the Board, Meta was right not to apply the newsworthiness allowance in this case, as the risk that the video could lead to additional and retaliatory violence had not passed when the



company reviewed it, given the overall context of widespread and ongoing gang and “self-defense” or “vigilante” violence in Haiti. The Board also finds that, to meet its human-rights responsibilities, Meta must ensure that moderation of content in Haiti, during this period of heightened risk, is effective and timely. The Board recommends Meta assess the timeliness and effectiveness of its responses to content escalated through the Trusted Partner program, including how effective Meta is in providing timely responses to escalations and what corrective measures Meta plans to adopt to improve response times to Trusted Partner escalations.

2: Case Description and Background

In May 2023, a Facebook user posted a video with a caption in Haitian Creole. The video shows a large group of people, who are wearing civilian clothing, walking into a police station and approaching a locked cell that has a man inside. According to Meta, the man inside the cell is a suspected member of the “5 Seconds Gang,” a well-armed and prominent gang in Haiti. The video also shows an individual from the group in the station attempting to break the cell’s lock. Several other people shout words of encouragement, including “we’re going to break the lock” and “they’re already dead.” Toward the end of the video, someone yells “bwa kale na boudaw.” According to Meta’s interpretation when referring the case to the Board, this phrase interpreted literally means “wooden stick up your ass”, and given the context indicated a call for the group “to take action against the person ‘bwa kale style’ – in other words, to lynch him.” Meta interprets the use of the term “bwa kale” to refer to the civilian movement of the same name, which involves civilians taking justice into their own hands against alleged gang members.

The video is accompanied by a caption describing what happens and stating that the “police cannot do anything, things are going to get weird.” According to linguistic experts consulted by the Board, the caption conveys a loss of faith in the police and a bleak outlook on what could happen next. The post was viewed over 500,000 times and the video was viewed around 200,000 times.



A Trusted Partner flagged the video to Meta as potentially violating 11 days after it was posted to Facebook, warning the content might incite further violence. Meta assessed the content and removed it from Facebook for violating its [Violence and Incitement](#) Community Standard. Meta’s Trusted Partner program is a network of non-governmental organizations, humanitarian agencies, human rights defenders and researchers from 113 countries around the world. Meta told the Board that the “greater the level of risk [of violence in a country], the higher the priority for developing relationships with Trusted Partners.” Trusted Partners can report content to Meta and provide feedback on the company’s content policies and enforcement. In this case, eight days after the Trusted Partner’s report, Meta determined the video included both a statement of intent to commit and a call for high-severity violence, and removed the content.

The following context is relevant to the Board’s decision. Haiti is experiencing “[unprecedented insecurity](#),” with gangs taking control of territory and terrorizing the population. Police are unable to address the violence and, in some cases, are [reported to be complicit](#). According to the [UN Special Representative to Haiti](#), “during the first quarter of the year, 1,647 criminal incidents – homicides, rapes, kidnappings and lynching – were recorded,” which is more than double the number compared with the same period in 2022. This rise in violence is taking place amid a political and humanitarian crisis. Haiti has not had an elected government since the assassination of President Jovenel Moïse in 2021 and has endured an ongoing cholera epidemic and [natural disasters](#). In March 2023, [Médecins Sans Frontières](#) (MSF) reported having to close one of its hospitals as a result of the intense violence in the country’s capital. Acting [Prime Minister Ariel Henry](#) has repeatedly appealed to the international community to send multinational forces to fight gang control, citing this as a necessary first step in “creating an environment for the State to function again.”

A civilian movement, referred to as “Bwa Kale,” has emerged in response to the rise in violence and the inability of the government or the police to protect the



population. A widely reported event that took place on April 24, 2023, has proven a pivotal moment for the movement. When Haitian police stopped a bus carrying 14 men with weapons, who were allegedly on their way to join an allied gang in a nearby district, a crowd gathered at the scene. Police stood back, and some were seen to help, as the crowd stoned the alleged gang members and burned them to death. Recordings of this event circulated widely on social media. According to [a report](#) by the National Human Rights Defense Network in Haiti, following the circulation of these recordings on social media, others, “armed with firearms, machetes, and tires, began to search for armed bandits, their relatives, or anyone suspected of having links with them, in order to lynch them.” According to the [UN High Commissioner for Human Rights](#), between April 24 and mid-August, “more than 350 people have been lynched by local people and vigilante groups.” In retaliation, gangs have taken revenge on those believed to be in or sympathetic to the movement.

On October 2, 2023, the United Nations Security Council [authorized](#) a year-long multinational security mission to Haiti. According to [reporting](#), it will be several months before forces are dispatched to Haiti.

3: Oversight Board Authority and Scope

The Board has authority to review decisions that Meta submits for review (Charter Article 2, Section 1; Bylaws Article 2, Section 2.1.1).

The Board may uphold or overturn Meta’s decision (Charter Article 3, Section 5), and this decision is binding on the company (Charter Article 4). Meta must also assess the feasibility of applying its decision in respect to identical content with parallel context (Charter Article 4). The Board’s decisions may include non-binding recommendations that Meta must respond to (Charter Article 3, Section 4; Article 4). When Meta commits to act on recommendations, the Board monitors their implementation.

4: Sources of Authority and Guidance



The following standards and precedents informed the Board’s analysis in this case:

I. *Oversight Board Decisions*

The most relevant previous decisions of the Oversight Board include:

- [Communal Violence in Indian state of Odisha](#)
- [Cambodian Prime Minister](#)
- [Brazilian General’s Speech](#)
- [Russian Poem](#)
- [Mention of the Taliban in News Reporting](#)
- [Shared Al Jazeera Post](#)
- [Öcalan's Isolation](#)
- [Former President Trump’s Suspension](#)
- [Claimed COVID Cure](#)
- [Nazi Quote](#)

II. *Meta’s Content Policies*

Meta’s [Violence and Incitement](#) policy “aims to prevent potential offline harm that may be related to content on Facebook.” The policy rationale notes that not all calls for violence are literal and likely to incite violence, therefore the company tries to “consider the language and context in order to distinguish casual statements from content that constitutes a credible threat to public or personal safety.” The policy rules prohibit “[s]tatements of intent to commit high-severity violence” and “[c]alls for high-severity violence.” Meta defines high-severity violence as a threat that could lead to death or is likely to be lethal. As part of the policy rationale, Meta explains that the company “see[s] aspirational or conditional threats directed at terrorists and other violent actors (e.g. ‘Terrorists deserve to be killed’), and [it] deem[s] those non-credible, absent specific evidence to the contrary.”



The Board’s analysis was informed by Meta’s [commitment to voice](#), which the company describes as “paramount,” and its value of safety. In explaining its commitment to voice, Meta states that “in some cases, we allow content – which would otherwise go against our standards – if it’s newsworthy and in the public interest.” This is known as the [newsworthiness allowance](#). It is a general policy exception applicable to all Community Standards. To potentially apply the allowance, Meta conducts a balancing test, assessing the public interest in the content against the risk of harm. Meta removes content “even if it has some degree of newsworthiness, when leaving it up presents a risk of harm, such as physical, emotional or financial harm, or direct threat to public safety.”

III. Meta’s Human-Rights Responsibilities

The UN Guiding Principles on Business and Human Rights (UNGPs), endorsed by the UN Human Rights Council in 2011, establish a voluntary framework for the human-rights responsibilities of private businesses. In 2021, Meta [announced](#) its [Corporate Human Rights Policy](#), in which it reaffirmed its commitment to respecting human rights in accordance with the UNGPs.

The Board’s analysis of Meta’s human-rights responsibilities in this case was informed by the following international standards:

- The rights to freedom of opinion and expression: Article 19, International Covenant on Civil and Political Rights ([ICCPR](#)), [General Comment No. 34](#), Human Rights Committee, 2011; UN Special Rapporteur on freedom of opinion and expression, reports: [A/HRC/38/35](#) (2018) and [A/74/486](#) (2019).
- The right to life: Article 6, ICCPR.
- The prohibition of advocacy of hatred that constitutes incitement to discrimination, hostility or violence: Article 20, para. 2, ICCPR; Rabat Plan of Action, UN High Commissioner for Human Rights report: [A/HRC/22/17/Add.4](#) (2013).



5: User Submissions

Following Meta’s referral and the Board’s decision to accept the case, the user was sent a message notifying them of the Board’s review and providing them with an opportunity to submit a statement to the Board. The user did not submit a statement.

6: Meta’s Submissions

Meta determined that the video constituted both a statement of intent to commit high-severity violence and a call for high-severity violence against the man being held in the cell, who, according to Meta, is a suspected member of the “5 Seconds Gang.” The “5 Seconds Gang” is a prominent gang in Haiti, so called because of “the perception that members will kill a person in that amount of time.” A member of the crowd can be heard on the video saying, “We’re going to break the lock...They’re already dead,” which Meta considered a statement of intent to kill the man. Meta also interpreted the phrase “bwa kale na boudaw” as a call to kill the man. Meta provided a broad analysis of the political, security and humanitarian situation in Haiti as background on the risk of harm posed by the content in question. Meta also noted that gang violence has become endemic in the country as government officials struggle to maintain authority and that “vigilantism is contributing to a culture of extrajudicial retributive violence.”

Meta considered two specific exceptions to the Community Standards as well as the newsworthiness allowance as part of its analysis. According to Meta, the company will allow content that violates the Violence and Incitement policy if it is “shared for the purpose of condemning or raising awareness of violence. The onus is on the user to make clear that one of those purposes is the intent.” In this case, Meta did not find a clear intent to condemn or raise awareness in the post. According to Meta, the fact the video was shared on a Facebook page that describes itself as a media page is not sufficient to satisfy this exception.



Meta also stated that it sometimes allows calls for high-severity violence in content that targets a person or entity designated under Meta’s [Dangerous Organizations and Individuals](#) (DOI) policy. According to Meta, this exception applies only if the company has confirmed that the target is a dangerous organization or individual, or a member of one. Meta informed the Board that the company has designated the “5 Seconds Gang” a dangerous organization. However, the company was unable to confirm the man in the cell shown in the video is a member of the gang. Had Meta been able to confirm his membership, then the content would not have violated the prohibition on call to action, according to the company.

Finally, in considering whether to apply the newsworthiness allowance, Meta determined the risk of harm from the post outweighed its public-interest value. Meta found that the video could contribute to violence either against the “5 Seconds Gang” or the Bwa Kale movement. While the content did have value in notifying others of impending violence and unfolding events, according to Meta, that value was diminished given the widespread coverage of the Bwa Kale movement.

Meta looked to the UN Rabat Plan of Action’s factors in assessing whether the post constitutes an incitement to violence and concluded that “the speech constituted an incitement to imminent violence” as the threat was “specific and connected to ongoing violent events.”

In response to the Board’s questions, Meta informed the Board that the company did not designate the situation in Haiti as a crisis under the Crisis Policy Protocol (CPP) as the company already had mitigation measures in place when the protocol was launched in August 2022.

The Board asked Meta 18 questions in writing. Questions related to Meta’s language capacity in enforcing its Community Standards in Haiti; processes for the review of reports from Trusted Partners and how the program relates to other systems Meta employs in crisis situations; and whether and how Meta used the Crisis Policy Protocol in Haiti. Meta answered all questions.



7: Public Comments

The Oversight Board received nine public comments. Seven of the comments were submitted from the United States and Canada, one from Asia Pacific and Oceania, and one from Europe.

To read public comments submitted for this case, please click [here](#).

8: Oversight Board Analysis

The Board examined whether this content should be removed by analyzing Meta’s content policies, human-rights responsibilities and values. The Board also assessed the implications of this case for Meta’s broader approach to content governance.

The Board selected this case to examine the role of social media in the context of extreme insecurity and violence, and how Meta’s policies and enforcement systems address content shared during an ongoing crisis. This case falls into the Board’s strategic priority of Crisis and Conflict Situations.

8.1 Compliance With Meta’s Content Policies

The Board finds that the content in this case violates the Violence and Incitement Community Standard. Nonetheless, the majority of the Board disagrees with Meta’s assessment on the application of the newsworthiness allowance. For the majority, given the delay of nearly three weeks in enforcement, Meta should have applied the newsworthiness allowance to allow the content to remain on Facebook at the time Meta reviewed the content.

I. Content Rules

a. Violence and Incitement

Meta prohibits “[s]tatements of intent to commit high-severity violence” and “[c]alls for high-severity violence.” The Board finds the content in this case



violates both policy lines. The content depicts incitement to violence in a context where there is a credible threat of offline harm to the person in the cell as well as others. The video shows a crowd of people as they attempt to break into a cell that holds a man who is alleged to be a gang member. People from the crowd shout that they will break in and that the man is “already dead.” These are statements that show intent to use lethal force. A member of the crowd shouts “bwa kale na boudaw,” a phrase that, in the [context in Haiti](#), constitutes a call to high-severity violence. While “bwa kale” has been used in various contexts, including in music and political messaging, in this case, the phrase is used in a context that mirrors deadly events in which civilians have killed suspected gang members or their allies.

Meta allows content that violates the Violence and Incitement policy to remain on the platform if it is shared to “raise awareness of or to condemn violence.” These exceptions are not included in the public-facing language of the policy but are provided in the internal set of instructions for content moderators. For the exception to apply, the company requires that the user make it clear that they are posting the content for either of the two reasons.

The Board finds the user in this case did not meet this burden; therefore, the content does not benefit from this exception as it is defined by Meta. The caption accompanying the video is descriptive and concludes with a statement that “[t]he police cannot do anything, things are going to get weird.” Describing the video or providing a neutral or ambiguous caption does not meet the standard established by Meta.

Meta also sometimes allows calls for high-severity violence when the target is a member of a designated [Dangerous Organization or Individual](#). This exception is referred to in Meta’s policy rationale for the Violence and Incitement policy, although it is not set out in the rules. The Board agrees that this exception does not apply in this case. However, the Board notes a number of concerns with this exception. First, the exception is not clearly articulated in the public-facing Community Standard. Second, as the list of individuals and organizations



designated under Meta’s policies is not public, there is no way for a user to know how this exception would apply to their content. The Board has repeatedly recommended that Meta should provide greater clarity and transparency on the Dangerous Organizations and Individuals policy (see [Mention of the Taliban in News Reporting](#); [Shared Al Jazeera Post](#); [Öcalan's Isolation](#); and [Nazi Quote](#)). Finally, according to Meta, the credibility of the threat is not a consideration in applying this exception. If the target is a designated entity or a violent actor, the content is deemed non-violating. The Board finds it troubling that credible threats against anyone designated under the opaque Dangerous Organizations and Individuals policy are exempted from the Violence and Incitement Community Standard.

b. *Newsworthiness Allowance*

While the Board finds the content violates the Violence and Incitement Community Standard, the majority of the Board disagrees with Meta on the application of the newsworthiness allowance in this case. First, the Board notes that the risk of harm and public interest involved in the newsworthiness analysis should be assessed at the time Meta is considering issuing the allowance, rather than at the time the user posted the content. Meta should update the public-facing language on the newsworthiness allowance to make this clear to users. Ideally, the two points in time should be close enough to avoid a different outcome, particularly in the context of widespread and escalating violence overcoming an entire nation. Unfortunately, in this case, nearly three weeks passed between the user posting the video and Meta’s removal of the content.

The majority of the Board finds that the risk of harm had significantly diminished when Meta made its decision (i.e., nearly three weeks after the incitement depicted in the video was posted) and Meta should have kept up the content by applying the allowance. The video has the potential to inform the public in Haiti, as well as abroad, of the realities of violence and the breakdown in public order at a time when Haiti is seeking international aid and



intervention. Whatever risk the content posed, including to identifiable individuals in the video, it had significantly diminished by the time Meta issued the allowance, as discussed further in section 8.2 (iii) analysis below. Had Meta reviewed the content soon after it was first posted, the risk of harm would have outweighed the public interest of the post, as in the [Communal Violence in Indian State of Odisha](#) case. In that case, Meta identified and removed the content within days of it being posted, at a time of heightened tensions and ongoing violence, when it posed a serious and likely risk of furthering violence, which outweighed the public interest value of the content. In this case, given Meta’s delay in reviewing the content, the risk of harm had significantly diminished, and was outweighed by its public interest value to safeguard access to information in order to inform the broader public of the situation in Haiti during this period. By the time Meta made its newsworthiness assessment, the post had been viewed 500,000 times and whatever risk of harm the video posed, had likely already materialized. As newsworthiness is assessed on escalation by Meta’s internal teams, Meta has the resources and expertise to make an even more context-sensitive assessment and to account for the change in circumstances when making that determination.

For a minority of Board Members, Meta was right not to apply the newsworthiness allowance in this case. While the risk of harm to the individuals depicted in the video was most acute in the days following the posting of the content, the risk that the video could lead to additional and retaliatory violence had not passed when Meta reviewed the content, given the overall context of widespread and ongoing violence and insecurity in Haiti. Therefore, the harms inherent in having the content on the platform still outweighed the public interest in publicizing the speech, as discussed further in section 8.2 (iii) below. The risk that others, upon seeing this video, could take up arms and join the movement and seek to punish someone had not abated. Neither had the possibility passed of a member of the “5 Seconds Gang,” or an affiliated gang, recognizing someone in the video and seeking revenge on them, on other members of the Bwa Kale movement or on members of the police force. For



these Board Members, the fact that several individuals are identifiable in the video and the risk of retaliation is well established and ongoing, means the content should not benefit from the allowance, even with the delay.

8.2 Compliance With Meta’s Human-Rights Responsibilities

The majority of the Board finds removing this content, three weeks after it was posted, was not necessary and proportionate and restoring the post to Facebook is consistent with Meta’s human-rights responsibilities. The Board also finds that, to meet its human-rights responsibilities, Meta must ensure that moderation of content in Haiti during this period of heightened risk is effective and timely.

Freedom of Expression (Article 19 ICCPR)

Article 19 of the ICCPR provides for broad protection of expression, including “commentary on one’s own and on public affairs” as well as expression that people may find offensive (General Comment 34, para 11). When restrictions on expression are imposed by a state, they must meet the requirements of legality, legitimate aim and necessity and proportionality (Article 19, para. 3, ICCPR). These requirements are often referred to as the “three-part test.” The Board uses this framework to interpret Meta’s voluntary human-rights commitments, both in relation to the individual content decision under review and what this says about Meta’s broader approach to content governance. As the UN Special Rapporteur on freedom of expression has stated, although “companies do not have the obligations of Governments, their impact is of a sort that requires them to assess the same kind of questions about protecting their users’ right to freedom of expression” ([A/74/486](#), para. 41).

1. Legality (Clarity and Accessibility of the Rules)

The principle of legality requires rules limiting expression to be accessible and clear, both to those enforcing the rules and those impacted by them (General Comment No. 34, para. 25). Rules restricting expression “may not confer



unfettered discretion for the restriction of freedom of expression on those charged with [their] execution” and must “provide sufficient guidance to those charged with their execution to enable them to ascertain what sorts of expression are properly restricted and what sorts are not” (*Ibid*). Applied to rules that govern online speech, the UN Special Rapporteur on freedom of expression has said they should be clear and specific ([A/HRC/38/35](#), para. 46). People using Meta’s platforms should be able to access and understand the rules and content reviewers should have clear guidance regarding their enforcement.

The Board finds that, as applied to the facts of this case, Meta’s prohibition of statements of intent to commit and calls for high-severity violence are clearly stated. The Board considers that the policy and its purpose, as applied to this case, are sufficiently clear to satisfy the legality requirement.

However, the Board notes that the “raising awareness or condemning violence” exception to the Violence and Incitement policy is still not available in the public-facing language of the policy. Failing to include these exceptions in the public-facing language of the Community Standard, and to explain that the onus is on the user to make their intent clear, raises serious legality concerns (see section 8.1 (1)(a) above). In the [Russian Poem](#) case, the Board recommended that Meta add to the public-facing language of its Violence and Incitement Community Standard its interpretation of the policy that allows for content containing statements with “neutral reference to a potential outcome of an action or an advisory warning,” and content that “condemns or raises awareness of violent threat.” Meta committed to making this change but has not updated the Violence and Incitement Community Standard accordingly. The Board highlights this recommendation again and urges Meta to add this exception to the public-facing language of the Community Standard.

II. Legitimate Aim

Under Article 19, para. 3 of the ICCPR, expression may be restricted for a defined and limited list of reasons. In this case, the Board finds the Violence and



Incitement Community Standard’s prohibition of statements of intent and calls to commit high-severity violence serves the legitimate aim of protecting public order and respecting the rights of others.

III. Necessity and Proportionality

The principle of necessity and proportionality provides that any restrictions on freedom of expression “must be appropriate to achieve their protective function; they must be the least intrusive instrument amongst those which might achieve their protective function; [and] they must be proportionate to the interest to be protected” ([General Comment No. 34 , para. 34](#)). The Board has previously used the Rabat Plan factors to analyze the necessity and proportionality of removing content under the Violence and Incitement Community Standard when public safety was at issue (see [Brazilian General’s Speech](#) and [Cambodian Prime Minister](#)). In this case, the Board looked to the Rabat Factors to evaluate the necessity and proportionality of removing this content. The Board also considered the lengthy delay in Meta reviewing this content, and what this indicates for the company’s ability to meet its human-rights responsibilities in moderating content in Haiti.

The majority of the Board finds removing this content, nearly three weeks after it was posted, was no longer necessary. The majority considered the context in Haiti, the extent and reach of the post and the likelihood of harm given the delay between posting of the content and its removal. The risk a post presents depends on the context in which it is shared. That context changed when Meta failed to act and, as a result, the video already had 200,000 views by the time of review. For these Board Members, given the delay in Meta’s review of the content and the high number of views that had previously occurred, whatever risk the content posed had likely already materialized. A timely assessment from Meta about this post would have affected the necessity and proportionality analysis and warranted its removal as in the [Communal Violence in Indian State of Odisha](#) case, in which the removal occurred within



days of the content being posted. Given the delay in Meta’s enforcement, the majority believes removal was no longer necessary.

Additionally, in a situation of protracted widespread violence and a breakdown of government authority and public order, sharing information becomes even more important for allowing communities to react to important events affecting them. Experts consulted by the Board highlighted the fact that people in Haiti rely on information shared on WhatsApp to stay informed of potential risks. In [a context](#) where “work of journalists is constrained by threats and violence, [where attacks] on journalists occur frequently, and impunity for perpetrators is the norm”, preserving access to information on social media becomes even more important. Ensuring content documenting events is not removed unnecessarily can aid in efforts to inform the public, and to identify and hold accountable those inciting and carrying out violence in Haiti.

In [Claimed COVID-19 Cure](#), the Board emphasized that Meta should explain the range of options it has at its disposal in achieving legitimate aims (such as preventing harm) and articulate why the selected one is the least intrusive means. As noted in that decision, Meta should publicly demonstrate three things in determining its least intrusive means: (1) the public interest objective could not be addressed through measures that do not infringe on speech; (2) among the measures that infringe on speech, Facebook (sic) has selected the least intrusive measure; and (3) the selected measure actually helps achieve the goal and is not ineffective or counterproductive (A/74/486, para. 51-52).

In this case, for example, given the international community’s interest in assessing the situation in order to help the people of Haiti (as described above), Meta should publicly justify why measures such as geo-blocking would be insufficient to avert harm. Given nearly three weeks had elapsed before Meta reviewed the content, Meta should also explain why measures such as preventing engagement with the content or employing demotions would not have been sufficient to minimize the risk of harm at that point. Rather, Meta seems to ask the Board to assess necessity and proportionality solely within a



binary up/down box instead of considering the impacts of its full range of tools, as is required by a serious human-rights approach to content moderation.

For a minority of the Board, removing this content is necessary and proportionate, especially given the context in Haiti, the extent and reach of the post, and the likelihood of harm. The Board found that this video was posted during a period of heightened risk, with intensifying gang violence and the start of a civilian movement of “self-defense” or “vigilante” violence against suspected gang members. This movement has previously taken suspected gang members from police custody to kill them by stoning, beating and setting them on fire. According to the [UN High Commissioner for Human Rights](#), between April 24 and mid-August, “more than 350 people have been lynched by local people and vigilante groups. Those killed have included 310 alleged gang members, 46 members of the public and a police officer.” Videos of such events have circulated on social media and have been connected to others taking up arms to join and search for suspected gang members in order to kill them. Additionally, according to reports from the [UN High Commissioner for Human Rights](#), members of the municipal government and police forces believed to be sympathetic to local self-defense groups have been killed by gangs in retaliation, as well as people believed to be in the movement. The leader of the “5 Seconds Gang” has previously threatened retaliation, including murder, on social media. The post names the precinct and shows the face of the person trying to break into the cell, as well as the faces of multiple people in the crowd. This post was viewed over 500,000 times. Given these facts, the threat of violence from this video circulating on Facebook was direct and imminent (General Comment 34, para 35), especially in the immediate period following its publication but also when Meta conducted its review. Additionally, for the minority, no measure, short of removal, would be sufficient to protect those depicted and those at risk of further violence spurred on by this video.

The Board is concerned about Meta's ability to proactively identify and effectively moderate content in Haiti in a timely manner. The Board notes the heightened risk of content directly contributing to harm in a context in which



public order and government services are absent, and extrajudicial and decentralized killing has become the main tool in a fight for power and control.

In this case, there was a significant delay in Meta evaluating and removing the content. This delay appears to be a result of the company's failure to invest adequate resources into moderating content in Haiti. The Board has previously raised concerns about the company's lack of investment in moderating content in non-English languages (see e.g. [Mention of the Taliban in News Reporting](#), [Shared Al Jazeera Post](#) and [Ocalan's Isolation](#)). In this case, Meta was not able to provide a timely assessment of a report from a Trusted Partner, which is one of the main tools Meta relies on in Haiti to identify potentially violating content. A [recent report](#) by one of Meta's Trusted Partners that evaluated the program found significant irregularity in response times from Meta and concluded that the program is under-resourced. Trusted Partners invest their time and resources to alert Meta of potentially dangerous content on its platforms. The Board is concerned that Meta is not resourcing its internal teams adequately enough to evaluate these reports in a timely manner.

Finally, Meta failed to activate its Crisis Policy Protocol in Haiti. In the [Former President Trump's Suspension](#) case, the Board urged Meta to develop and publish a policy to govern its responses to crises and novel situations where its regular processes would not prevent or avoid imminent harm. In response, Meta created the Crisis Policy Protocol, which aims to "codify [the company's] policy-specific responses to ensure [Meta] is timely, systematic and proportionate in a crisis" ([Crisis Policy Protocol](#), Policy Forum Minutes, January 25, 2022). In this case, Meta told the Board that the company did not designate the situation in Haiti as a crisis under the protocol as it is "designed to facilitate timely assessment and mitigation of novel or emergent crises," and the company already had risk-mitigation measures in place in Haiti when the Crisis Policy Protocol came into use in August 2022. However, the Board is concerned that if the company fails to use the Crisis Policy Protocol in such situations, it will fail to deliver principled and timely moderation in these circumstances. Many crises and conflicts around the world are ongoing or have periods of acute



violence or harm that subside and re-emerge depending on the circumstances. Meta must have a mechanism in place to assess risks in such crises and transition from existing mitigation measures to those provided by the Crisis Policy Protocol. Failure to use the Crisis Policy Protocol under such circumstances undermines the company's and the public's ability to assess the effectiveness of the protocol in meeting its aims.

The Board understands that Meta must make difficult decisions when it comes to how it prioritizes resourcing for its various content-moderation systems (i.e. developing language-specific classifiers, hiring content moderators, deploying the Crisis Policy Protocol or prioritizing operational measures such as Trusted Partners). However, to meet its human-rights responsibilities, Meta must ensure that moderation of content in Haiti, during this period of heightened risk, is effective and timely.

9: Oversight Board Decision

The Oversight Board overturns Meta's decision to take down the content, requiring the post to be restored.

10: Recommendations

Enforcement

1. To address the risk of harm, particularly where Meta has no or limited proactive moderation tools, processes or measures to identify and assess content, Meta should assess the timeliness and effectiveness of its responses to content escalated through the Trusted Partner program.

The Board will consider this recommendation implemented when Meta both shares the results of this assessment with the Board – including the distribution of average time to final resolution for escalations originating from Trusted Partners disaggregated by country, Meta's own internal goals for time to final resolution, and any corrective measures it is taking in case those targets are not met – as well as



publishes a public-facing summary of its findings to demonstrate it has complied with this recommendation.

Policy

The Board also reiterates the following recommendation from the Russian Poem case:

Meta should add to the public-facing language of its Violence and Incitement Community Standard that it interprets the policy to allow content containing statements with “neutral reference to a potential outcome of an action or an advisory warning,” and content that “condemns or raises awareness of violent threat.”

***Procedural Note:**

The Oversight Board’s decisions are prepared by panels of five Members and approved by the majority of the Board. Board decisions do not necessarily represent the personal views of all Members.

For this case decision, independent research was commissioned on behalf of the Board. The Board was assisted by an independent research institute headquartered at the University of Gothenburg, which draws on a team of over 50 social scientists on six continents, as well as more than 3,200 country experts from around the world. The Board was also assisted by Duco Advisors, an advisory firm focusing on the intersection of geopolitics, trust and safety, and technology. Memetica, an organization that engages in open-source research on social media trends, also provided analysis. Linguistic expertise was provided by Lionbridge Technologies, LLC, whose specialists are fluent in more than 350 languages and work from 5,000 cities across the world.